Understanding high-order correlations using a synergy-based decomposition of the total entropy

Fernando Rosas*, Vasilis Ntranos†, Christopher J. Ellison‡,
Marian Verhelst* and Sofie Pollin*

* Departement Elektrotechniek, KU Leuven
† Communication Sciences Institute, University of Southern California
‡ Center for Complexity and Collective Computation, University of Wisconsin-Madison
fernando.rosas@esat.kuleuven.be

Abstract

The interactions between three or more variables are frequently nontrivial, poorly understood, and yet, are paramount for future advances in fields such as multiuser information theory, neuroscience, and genetics. We introduce a novel framework that characterizes the ways in which random variables can share information, based on the notion of information synergy. The framework is then applied to several network information theory problems, providing a more intuitive understanding of their fundamental limits.

1 Introduction

Developing a framework for understanding the correlations that can exist between multiple signals is crucial for the design of efficient and distributed communication systems. For example, consider a network that measures the weather conditions (e.g. temperature, humidity, etc) in a specific region. Given the nature of the underlying processes being measured, one should expect that the sensors will generate strongly correlated data. A haphazard design will not account for these correlations and, undesirably, will process and transmit redundant information across the network.

Higher-order correlations are also of more general interest. In neuroscience, researchers desire to identify how various neurons affect an organism's overall behavior, asking to what extent the different neurons are providing redundant or synergistic signals [1]. In genetics, the interactions and roles of multiple genes with respect to phenotypic phenomena are studied, e.g. by comparing results from single and double knockout experiments [2].

In this work we propose a new framework for understanding complex correlations, which is novel in combining the notion of hierarchical decomposition as developed in [3], with the notion of information synergy as proposed in [4]. In contrast to [3], we focus on the dual total correlation instead of the total correlation, which is more directly related to the shared information within the system. In contrast to [4], we analize the joint entropy instead of the mutual information. Our framework provides new insight to various problems of Network Information Theory. Interestingly, many of the problems of Network Information Theory that have been solved are related to systems which present a simple structure in terms of synergies and redundancies, while most of the open problems possess a more complex mixture of them.

In the following, Section 2 introduces the notions of hierarchical decomposition of correlations and synergistic information, providing the necessary background for an unfamiliar reader. Then, Section 3 presents our decomposition for the joint entropy, focusing on the case of three variables and leaving its generalization for a future work. Section 4 applies this framework in settings of fundamental importance for Network Information Theory. Finally, Section 5 summarizes our main conclusions.

2 Preliminaries

One way of analyzing the interactions between the random variables $\mathbf{X} = (X_1, \ldots, X_N)$ is to study the matrix properties of $\mathcal{R}_{\mathbf{X}} = \mathbb{E}\{\mathbf{X}\mathbf{X}^t\}$. However, this only captures linear relationships and hence the picture provided by $\mathcal{R}_{\mathbf{X}}$ is incomplete. Another possibility is to study the matrix $\mathcal{I}_{\mathbf{X}} = [I(X_i; X_j)]_{i,j}$ of mutual informations. This matrix captures the existence of both linear and nonlinear dependencies, but its scope is restricted to pairwise relationships and thus, it misses all higher-order structure. To see how this can happen, consider two independent fair coins X_1 and X_2 and let $X_3 := X_1 \oplus X_2$ be output of an `XOR` logic gate. The mutual information matrix $\mathcal{I}_{\mathbf{X}}$ has all off-diagonal elements equal to zero, making it indistinguishable from an alternative situation where X_3 is another independent fair coin.

For the case of $\mathcal{R}_{\mathbf{X}}$, the standard next step would be to consider higher order moment matrices such co-skewness and co-kurtosis. We seek their information-theoretic analogs, which complement the description provided by $\mathcal{I}_{\mathbf{X}}$. One method of doing this is by studying the information contained in marginal distributions of increasingly larger sizes; this approach is presented in Section 2.1. Other methods try to provide a direct representation of the information that is shared between the various random variables; they are discussed in Section 2.2.

2.1 Negentropy and total correlation

When the random variables that compose a system are independent, their joint distribution is given by the product of their marginal distributions. Hence, in this case the marginals contain all that is to be learned about the statistics of the entire system. However, arbitrary joint p.d.f.s can contain information that is not present in their marginals. To quantify this idea, let us consider N discrete random variables $\mathbf{X} = (X_1, \ldots, X_N)$ with joint p.d.f. $p_{\mathbf{X}}$, where each X_j takes values in a finite set with cardinality Ω_j. The maximal amount of information that could be stored in any such system is $H^{(1)} = \sum_j \log \Omega_j$, which corresponds to the entropy of the p.d.f. $p_{\mathbf{U}} := \prod_j \overline{p}_{X_j}$, where $\overline{p}_{X_j}(x) = 1/\Omega_j$ is the uniform distribution for each random variable X_j. On the other hand, the joint entropy $H(\mathbf{X})$ with respect to the true distribution $p_{\mathbf{X}}$ measures the actual uncertainty that the system possesses. Therefore, the difference $\mathcal{N}(\mathbf{X}) := H^{(1)} - H(\mathbf{X})$ corresponds to the decrease of the uncertainty about the system that occurs when one learns its p.d.f. – i.e. the information about the system that is contained in its statistics. This quantity is known as *negentropy* [5], and can be also computed as

$$\mathcal{N}(\mathbf{X}) = D\big(\textstyle\prod_j p_{X_j} \,\big|\big|\, p_{\mathbf{U}}\big) + D\big(p_{\mathbf{X}} \,\big|\big|\, \textstyle\prod_j p_{X_j}\big) \ , \tag{1}$$

where p_{X_j} is the marginal of the variable X_j and $D(\cdot||\cdot)$ is the Kullback-Leibler divergence. In this way, (1) decomposes the negentropy into a term that corresponds to the information given by simple marginals and a term that corresponds to higher order marginals. The second term is known as the *Total Correlation* (TC) and has been suggested as an extension of the notion of mutual information for multiple variables.

An elegant framework for decomposing the TC can be found using the framework presented in [3]. Let us call k-marginals the distributions that are obtained by marginalizing the joint p.d.f. over $N - k$ variables. In the case where only the 1-marginals are known, the simplest guess for the joint distribution is $\tilde{p}^{(1)}_{\mathbf{X}} = \prod_j p_{X_j}$. One way of generalizing this for when the k-marginals are known is by using the *maximum entropy principle*, which suggests to choose the distribution that maximizes the joint entropy while satisfying the constrains given by the partial (k-marginal) knowledge. Let us denote by $\tilde{p}^{(k)}_{\mathbf{X}}$ the p.d.f. which achieves a maximum entropy while being consistent with the k-marginals, and let $H^{(k)} = H(\{\tilde{p}^{(k)}_{\mathbf{X}}\})$ denote its entropy. Then, it can be showed the following generalized Pythagorean relationship for the total correlation:

$$\mathrm{TC} = \sum_{k=2}^{N} D(\tilde{p}^{(k)}||\tilde{p}^{(k-1)}) = \sum_{k=2}^{N} \left(H^{(k-1)} - H^{(k)}\right) \triangleq \sum_{k=2}^{N} \Delta H^{(k)} \ . \tag{2}$$

Above, $\Delta H^{(k)} \geq 0$ measures the information that is provided by the k marginals and not by the $k-1$ ones. In general, information located in terms with high values of k correspond to complex correlations between many variables, which cannot be reduced to a combination of simpler correlations between smaller groups.

2.2 Yeung's decomposition and synergistic information

Another approach to study the correlations between many random variables is to analyze the way in which they share information, which can be done by decomposing the joint entropy of the system. For the case of two variables, the joint entropy can be decomposed as $H(X_1, X_2) = I(X_1; X_2) + H(X_1|X_2) + H(X_2|X_1)$, suggesting that it can be divided into shared information, $I(X_1; X_2)$, and informations that are exclusively located in just one variable, $H(X_1|X_2)$ and $H(X_2|X_1)$. In systems with more than two variables, one can still compute the information that is exclusively located in one element as $H_{(1)} := \sum_j H(X_j|\mathbf{X}_j^c)$, where $\mathbf{X}_j^c$ denote all the system variables except X_j. The difference between the joint entropy and the sum of informations contained in just one location defines the *Dual Total Correlation* (DTC),

$$\mathrm{DTC} = H(\mathbf{X}) - H_{(1)}, \tag{3}$$

which measures the portion of the joint entropy that is shared between two or more variables of the system. As in (2), it would be appealing to look for a decomposition of the DTC of the form $\mathrm{DTC} = \sum_{k=2}^{N} \Delta H_{(k)}$, where $\Delta H_{(k)} \geq 0$ would measure the information that is shared by k variables.

One possible decomposition for the DTC is provided by the *I-measure* [6]. For the case of three variables, this decomposition can be written as

$$\mathrm{DTC}_{N=3} = [I(X_1; X_2|X_3) + I(X_2; X_3|X_1) + I(X_3; X_1|X_2)] + I(X_1; X_2; X_3) \ . \tag{4}$$

The last term is known as the *co-information* [7] and can be calculated as $I(X_1; X_2; X_3) = I(X_1; X_2) - I(X_1; X_2|X_3)$, being other candidate for extending the mutual information to multiple variables. Although it is tempting to associate the term in square brackets of (4) with $\Delta H_{(2)}$ and the co-information with $\Delta H_{(3)}$, this would not be very intuitive since the co-information can be negative. Conventionally, we think of the conditional mutual information as the information contained in X_1 and X_2 that is not contained in X_3, but this quantity should be strictly less than the total information shared by X_1 and X_2. The counterintuitive fact that sometimes $I(X_1; X_2) \leq I(X_1; X_2|X_3)$ suggests that the conditional mutual information can capture information that extends beyond X_1 and X_2, incorporating higher-order effects with X_3.

An extended treatment of the conditional mutual information and it relationship with the mutual information can be found in [4]. For presenting those ideas, let us consider two random variables X_1 and X_2 which are used to predict X_3. The total predictability, i.e., the information X_1 and X_2 provide about X_3, is given by $I(X_1, X_2; X_3) = I(X_1; X_3) + I(X_2; X_3|X_1)$. Is natural to think that the information provided by X_1, $I(X_1; X_3)$, can be unique or redundant with respect of the information provided by X_2. On the other hand, $I(X_2; X_3|X_1)$ must contain the unique contribution of X_2. However, the fact that $I(X_2; X_3|X_1)$ can be larger than $I(X_2; X_3)$ (while the latter contains both the unique and redundant contributions of X_2) suggests that there can be an additional predictability that is accounted only by the conditional mutual information. This predictability, which is not contained in any single predictor but is only revealed by both X_1 and X_2, is called *synergistic mutual information*. As an example of this, consider again the case in which X_1 and X_2 are independent random bits and $X_3 = X_1 \oplus X_2$. Then, it can be seen that $I(X_1; X_3) = I(X_2; X_3) = 0$ but $I(X_1, X_2; X_3) = 1$. Hence, neither X_1 nor X_2 individually provide information about X_3, although together they fully determine it.

Further discussions about the notion of information synergy can be found in [8–11].

3 A non-negative joint entropy decomposition

In this section we present our non-negative decomposition of the joint entropy, which is based on the notion of information synergy. It is important to note that there is an ongoing debate about the best way of characterizing and computing the synergy in arbitrary systems, as the commonly used axioms are not enough for specifying a unique formula [9]. Nevertheless, our approach in this work is to explore how far one can reach based only on the axioms. In this way, our results are going to be consistent to any choice of formula that is consistent with the axioms.

In the following, Section 3.1 presents the axioms of Information Synergy that are used in this work. Then, Section 3.2 will first present the decomposition for an arbitrary system of three variables. Sections 3.2.1 and 3.2.2 specify the decomposition for the important cases of Markov chains and pairwise independent predictors, which provide the basis for the applications explored in Section 4.

3.1 Information synergy axioms

We proceed to determine a number of desired properties that a decomposition of the mutual information should possess. Note that we initially privilege X_3, but our decomposition will end up being symmetric in each random variables.

Definition A decomposition of the mutual information is provided by the functions $I_\cap(X_1X_2; X_3)$, $I_\mathrm{S}(X_1X_2; X_3)$ and $I_\mathrm{un}(X_1; X_3|X_2)$ which satisfy the following axioms:

(1) $I(X_1; X_3) = I_\cap(X_1X_2; X_3) + I_\mathrm{un}(X_1; X_3|X_2)$.

(2) $I(X_1; X_3|X_2) = I_\mathrm{un}(X_1; X_3|X_2) + I_\mathrm{S}(X_1X_2; X_3)$.

(3) *Weak symmetry*: $I_\cap(X_1X_2; X_3) = I_\cap(X_2X_1; X_3)$, $I_\mathrm{S}(X_1X_2; X_3) = I_\mathrm{S}(X_2X_1; X_3)$ and $I_\mathrm{un}(X_1; X_3|X_2) = I_\mathrm{un}(X_3; X_1|X_2)$.

(4) Non-negativity: $I_\cap(X_1X_2; X_3) \geq 0$, $I_\mathrm{S}(X_1X_2; X_3) \geq 0$, and $I_\mathrm{un}(X_1; X_3|X_2) \geq 0$.

Intuitively, $I_\cap(X_1X_2; X_3)$ measures the redundancy of X_1 and X_2 for predicting X_3, $I_\mathrm{un}(X_1; X_3|X_2)$ quantifies the unique information that is provided by X_1 (and not X_2) about X_3, and $I_\mathrm{S}(X_1X_2; X_3)$ is the synergistic mutual information between X_1 and X_2 about X_3. Note that the weak symmetry of the unique information is not strictly necessary for proving our results, but is adopted here because it allows for a more intuitive development of our ideas.

Using the symmetry of the mutual information and Axiom (1), we can show that

$$I_\cap(X_1X_2; X_3) + I_\mathrm{un}(X_1; X_3|X_2) = I(X_3; X_1) = I_\cap(X_3X_2; X_1) + I_\mathrm{un}(X_3; X_1|X_2) \quad (5)$$

Then, by using the weak symmetry of the unique information, it follows that the redundancy also satisfies *strong symmetry*, i.e. $I_\cap(X_1X_2; X_3) = I_\cap(X_3X_2; X_1)$. In a similar way, using the symmetry of the conditional entropy one can show that

$$I_\mathrm{un}(X_1; X_3|X_2) + I_\mathrm{S}(X_1X_2; X_3) = I(X_3; X_1|X_2) = I_\mathrm{un}(X_3; X_1|X_2) + I_\mathrm{S}(X_3X_2; X_1). \quad (6)$$

Using again the weak symmetry of the unique information, one can prove the strong symmetry of the synergy. In order to reflect the strong symmetry of these functions, we will henceforth denote the redundancy and synergy as $I_\cap(X_1;X_2;X_3)$ and $I_\mathrm{S}(X_1;X_2;X_3)$, respectively.

3.2 Decomposition for three variables

Inspired by the non-negative decomposition of the TC, our approach is to build a non-negative decomposition of the joint entropy which is based on a non-negative decomposition of the DTC. For the case of three variables, we let

$$H_{(1)} = H(X_1|X_2,X_3) + H(X_2|X_1,X_3) + H(X_3|X_1,X_2) \tag{7}$$

$$\Delta H_{(2)} = I_\mathrm{un}(X_1;X_2|X_3) + I_\mathrm{un}(X_2;X_3|X_1) + I_\mathrm{un}(X_3;X_1|X_2) \tag{8}$$

$$\Delta H_{(3)} = I_\cap(X_1;X_2;X_3) + 2I_\mathrm{S}(X_1;X_2;X_3) \tag{9}$$

and define the decomposition of the joint entropy as:

$$H(X_1,X_2,X_3) = H_{(1)} + \Delta H_{(2)} + \Delta H_{(3)}. \tag{10}$$

Comparing (10) with (3) yields DTC $= \Delta H_{(2)} + \Delta H_{(3)}$. Each $\Delta H_{(k)}$ term is non-negative because of Axiom (4), and hence (10) yields a non-negative decomposition of the joint entropy, where each of the corresponding terms captures the information that is shared by one, two or three variables.

In the following, we will analyze two scenarios for which explicit formulas for (8) and (9) can be found.

3.2.1 Markov chains

Let us consider the case in which $X_1 - X_2 - X_3$ form a Markov chain. Because of the conditional independence of X_1 and X_3 with respect to X_2 one has that $I(X_1;X_3|X_2) = 0$. Therefore, by using Axiom (2), it is clear that $I_\mathrm{un}(X_1;X_3|X_2) = 0$, which is consistent with the fact that X_1 and X_3 should not share information that is not also present in X_2. Moreover, using this and Axiom (1), one can find that the redundant information of the Markov chain is $I_\cap(X_1;X_2;X_3) = I(X_1;X_3)$. Using this and Axiom (1), one can show that $I_\mathrm{un}(X_1;X_2|X_3) = I(X_1;X_2) - I(X_1;X_3)$ and $I_\mathrm{un}(X_2;X_3|X_1) = I(X_2;X_3) - I(X_1;X_3)$. Therefore, the information that is shared by pairs of variables in a Markov chain can be found to be

$$\Delta H_{(2)} = I(X_1;X_2) + I(X_2;X_3) - 2I(X_1;X_3) \ . \tag{11}$$

Using again $I(X_1;X_3|X_2) = 0$ and Axiom (2), it is direct to see that $I_\mathrm{S}(X_1;X_2;X_3) = 0$. Therefore, in this case

$$\Delta H_{(3)} = I_\cap(X_1;X_2;X_3) = I(X_1;X_3). \tag{12}$$

3.2.2 Pairwise independent predictors (PIP)

Let us assume that X_1 and X_2 are pairwise independent, and therefore $I(X_1;X_2) = 0$. Then, using Axiom (1), it is direct to see that $I_\mathrm{un}(X_1;X_2|X_3) = I_\cap(X_1;X_2;X_3) = 0$, which in turn allows to show that $I_\mathrm{un}(X_1;X_3|X_2) = I(X_1;X_3)$ and $I_\mathrm{un}(X_2;X_3|X_1) = I(X_2;X_3)$. Therefore, in this case

$$\Delta H_{(2)} = I(X_1;X_3) + I(X_2;X_3) \ , \tag{13}$$

which shows that the positive mutual information terms correspond to information that is shared only by two variables. Using these results and Axiom (2), one can also compute the synergy directly as $I_\mathrm{S}(X_1;X_2;X_3) = I(X_1;X_3|X_2) - I(X_1;X_3) = I(X_1;X_2|X_3)$. Therefore, in this case we have

$$\Delta H_{(3)} = 2I(X_1;X_2|X_3), \tag{14}$$

which measures the correlations between X_1 and X_2 that are introduced by X_3.

4 Applications to Network Information Theory

In this section we will apply the framework presented in Section 3 to develop new intuitions over three fundamental scenarios in Network Information Theory [12]. In the following, Section 4.1 uses the general framework to analyze the Slepian-Wolf coding for three sources, which is a fundamental result in the literature of distributed source compression. Then, Section 4.2 applies the results for PIP to the multiple access channel (MAC), which is one of the fundamental settings in multiuser information theory. Finally, Section 4.3 applies the results for Markov chains to the wiretap channel, which constitutes one of the main models of information-theoretic secrecy.

4.1 Slepian-Wolf coding

The Slepian-Wolf coding gives lower bounds for the data rates that are required to transfer the information contained in various data sources. Let us denote as R_k the data rate of the k-th source and define $\Delta R_k = R_k - H(X_k|\mathbf{X}_k^{\mathrm{c}})$ as the extra data rate that each source has above what is needed for their own exclusive information (c.f. Section 2.2). Then, in the case of two sources X_1 and X_2, the well-known Slepian-Wolf bounds can be re-written as $\tilde{R}_1 \geq 0$, $\tilde{R}_2 \geq 0$, and $\tilde{R}_1 + \tilde{R}_2 \geq I(X_1;X_2)$. The last inequality states that $I(X_1;X_2)$ corresponds to shared information that can be transmitted by any of the two sources.

Let us consider now the case of three sources, and denote $R_S = I_{\mathrm{S}}(X_1;X_2;X_3)$. The Slepian-Wolf bounds provide seven inequalities, which can be re-written as

$$\tilde{R}_i \geq 0,\ i \in \{1,2,3\} \tag{15}$$

$$\tilde{R}_i + \tilde{R}_j \geq I_{\mathrm{ex}}(X_i;X_j|X_k) + R_S,\ i,j \in \{1,2,3\}, i<j \tag{16}$$

$$\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3 \geq \Delta H_{(2)} + \Delta H_{(3)} \tag{17}$$

Above, (17) states that all shared information (i.e. the DTC) needs to be accounted by the extra rate of the sources, and (16) that every pair needs to to take care of their unique information and the synergy. Note that, because of (9), the redundancy can be included in only one of the rates while the synergy has to be included in at least two.

4.2 Multiple Access Channel

Let us consider a multiple access channel (MAC), where two pairwise independent transmitters send X_1 and X_2 and a receiver gets X_3 as shown in Fig. 1, forming a PIP system (c.f. Section 3.2.2). It is well-known that, for a given distribution $(X_1,X_2) \sim p(x_1)p(x_2)$, the achievable rates R_1 and R_2 satisfy the capacity constrains $R_1 \leq I(X_1;X_3|X_2)$, $R_2 \leq I(X_2;X_3|X_1)$ and $R_1+R_2 \leq I(X_1,X_2;X_3)$.

As the transmitted random variables are pairwise independent, one can apply the results of Section 3.2.2. Hence, there is no redundancy and $I_{\mathrm{S}} = I(X_1;X_3|X_2) - I(X_1;X_3)$. Let us introduce a shorthand notation for the remaining three terms : $C_1 = I_{\mathrm{un}}(X_1;X_3|X_2) = I(X_1;X_3)$, $C_2 = I_{\mathrm{un}}(X_2;X_3|X_1) = I(X_2;X_3)$ and $C_S = I_{\mathrm{S}}(X_1;X_2;X_3)$. Then, one can re-write the bounds for the transmission rates as

$$R_1 \leq C_1 + C_S, \quad R_2 \leq C_2 + C_S \quad \text{and} \quad R_1 + R_2 \leq C_1 + C_2 + C_S. \tag{18}$$

From this, it is clear that while each transmitter has an unique portion of the channel with capacity C_1 or C_2, their interaction creates *synergistically* an additional capacity that is given by $C_S = I_{\mathrm{S}}(X_1;X_2;X_3)$.

There exists an interesting relationship between (18) and the bounds provided by Slepian-Wolf coding for two sources A and B. In effect, $H(A|B)$ and $H(B|A)$ correspond to exclusive information contents that needs to be transmitted by each source,

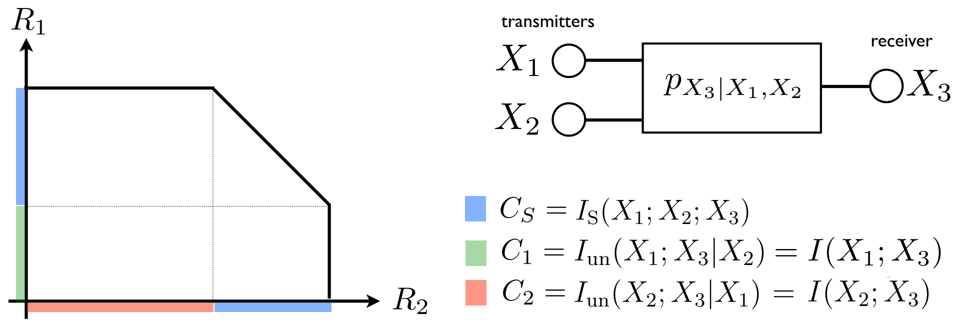

Figure 1: Multiple Access Channel

while C_1 and C_2 are the capacities of the unique portions of the channel that cannot be shared. Also, the mutual information $I(A; B)$ is the information that can be transmitted by either of the variables, while the synergetic capacity C_S corresponds to the part of the channel that can be shared between the users.

4.3 Degraded Wiretap Channel

Consider a communication system with an eavesdropper (shown in Fig. 2), where the transmitter sends X_1, the intended receiver gets X_2 and the eavesdropper receives X_3. For simplicity of the exposition, let us consider the case of a degraded channel where $X_1 - X_2 - X_3$ form a Markov chain. Using the results of Section 3.2.1, one can see that in this case there is no synergy but only redundancy and unique information between X_1 or X_3 with X_2.

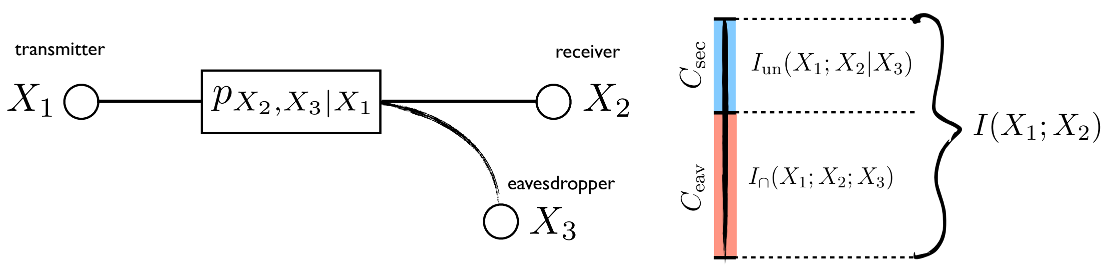

Figure 2: Wiretap Channel

In this scenario, it is known that for a given input distribution p_{X_1} the rate of secure communication that can be achieved is upper bounded by

$$C_{\text{sec}} = I(X_1; X_2) - I(X_1; X_3) = I_{\text{un}}(X_1; X_2|X_3), \tag{19}$$

which is precisely the unique information between X_1 and X_2. Also, as intuition would suggest, the eavesdropping capacity is equal to the redundancy and is given by

$$C_{\text{eav}} = I(X_1; X_2) - C_{\text{sect}} = I(X_1; X_3) = I_\cap(X_1; X_2; X_3). \tag{20}$$

5 Conclusions

We proposed a framework for understanding how multiple random variables can share information, based on a novel decomposition of the joint entropy. We showed how the axioms, on which our framework is based, allow us to find concrete expressions for all the terms of the decomposition for Markov chains and for the case where two variables are pairwise independent. These results allow for an intuitive understanding of the optimal information-theoretic strategies for several fundamental scenarios in Network Information Theory.

The key insight that this framework provides is that while there is only one way in which information can be shared between two random variables, it can be shared in two different ways between three: redundantly or synergistically. This important distinction has shed new light in the understanding of high-order correlations, whose consequences have only begun to be explored.

Acknowledgments

We want to thank David Krakauer and Jessica Flack for providing the inspiration for this research. We also thank Bryan Daniels, Vigil Griffith and Martin Ugarte for helpful discussions. This work was partially supported by a grant to the Santa Fe Institute for the study of complexity and by the U.S. Army Research Laboratory and the U.S. Army Research Office under contract number W911NF-13-1-0340. FR would also like to acknowledge the support of the F+ fellowship from KU Leuven and the SBO project SINS, funded by the Agency for Innovation by Science and Technology IWT, Belgium.

References

[1] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia, "Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies," *Neural Computation*, vol. 12, no. 11, pp. 2621–2653, 2000.

[2] D. Deutscher, I. Meilijson, S. Schuster, and E. Ruppin, "Can single knockouts accurately single out gene functions?" *BMC Systems Biology*, vol. 2, no. 1, p. 50, 2008.

[3] S.-I. Amari, "Information geometry on hierarchy of probability distributions," *Information Theory, IEEE Transactions on*, vol. 47, no. 5, pp. 1701–1711, Jul 2001.

[4] V. Griffith and C. Koch, "Quantifying synergistic mutual information," in *Guided Self-Organization: Inception*, ser. Emergence, Complexity and Computation, M. Prokopenko, Ed. Springer Berlin Heidelberg, 2014, vol. 9, pp. 159–190.

[5] L. Brillouin, "The negentropy principle of information," *Journal of Applied Physics*, vol. 24, no. 9, pp. 1152–1163, 1953.

[6] R. W. Yeung, "A new outlook on shannon's information measures," *Information Theory, IEEE Transactions on*, vol. 37, no. 3, pp. 466–474, 1991.

[7] A. J. Bell, "The co-information lattice," in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA*, vol. 2003. Citeseer, 2003.

[8] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multivariate information," *arXiv preprint arXiv:1004.2515*, 2010.

[9] V. Griffith, E. K. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield, "Intersection information based on common randomness," *Entropy*, vol. 16, no. 4, pp. 1985–2000, 2014.

[10] M. Harder, C. Salge, and D. Polani, "Bivariate measure of redundant information," *Physical Review E*, vol. 87, no. 1, p. 012130, 2013.

[11] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, "Quantifying unique information," *Entropy*, vol. 16, no. 4, pp. 2161–2183, 2014.

[12] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.